

Д.В. АПАНОВИЧ, М.П. КОНЦЕВОЙ

БрГУ имени А.С. Пушкина (г. Брест, Беларусь)

КОМПЬЮТЕРНОЕ МОДЕЛИРОВАНИЕ ПРОСОДИКИ В ЛИНГВИСТИЧЕСКОМ КОРПУСЕ ЗВУЧАЩЕЙ РЕЧИ

Корпус звучащей речи – структурированная совокупность оцифрованных речевых фрагментов с программными средствами поиска и метаданным об этих фрагментах. Речевых корпуса позволяют реализовать недоступную ранее возможность для проведения крупномасштабных и статистически достоверных исследований речи на аутентичном и репрезентативном материале. Просодическая разметка речевых корпусов моделирует лишь самые простые и регулярные слои просодии, ориентируясь на определенные исследовательские задачи. Важнейшая содержательная задача моделирования транскрипции корпуса звучащей речи – определение формальных критериев членения речевого потока на элементарные единицы. Если в семантико-синтаксическом подходе такими структурами служат клаузы (предикции), то просодический подход предполагает выделение интонационных единиц, организуя структуру вокруг одного главного акцента, имеющего единый интонационный контур. К таким структурам можно отнести оборванные единицы в случае нарушения в развертывании речи, постпозитивные уточняющие единицы, парентетические фрагменты, интонационно автономные регуляторные единицы, интонационно автономные обращения. слова *да*, *нет* с сопровождающими их клитиками, вынесенные топики, имеющие прономинальную копию или полнозначную кореферентную ИГ в составе следующей предикации. Реализация предложенной модели может быть осуществлена с помощью CLAN (Child Language Analysis) – кроссплатформенного приложения с открытым исходным кодом для создания и анализа транскриптов в базе данных Child Language Exchange System (CHILDES). CLAN включает в себя редактор создания связей между словами и высказываниями (в стенограмме, аудио- или видеоматериалах) в формате транскрипции анализа разговоров Conversation analysis (CA). В отдельных окнах вызываются все основные инструменты корпусного анализа (ключевые слова и строки, конкордация, подсчет частот, поиск регулярных выражений и др.), дополнительный программный инструментарий анализа паттернов непредвиденных ситуаций между речами (длины высказываний и слов, кластеров совпадений и др.), средства разметки специальных сегментов текста, редактор системы кодирования. Результаты работы данных приложений отправляются в основное окно вывода CLAN.